

R-Indicators for the 2019 Survey of Consumer Finances (SCF)

April 14, 2021

Katie Archambeau, NORC, University of Chicago
Kate Bachtell, NORC, University of Chicago
Cathy Haggerty, NORC, University of Chicago
Shannon Nelson, NORC, University of Chicago
Micah Sjoblom, NORC, University of Chicago
Steven Pedlow, NORC, University of Chicago
Kevin Moore, Board of Governors of the Federal Reserve System



Background of Adaptive Survey Design (ASD)

Adaptive Survey Design Definition and Goals

Adaptive/Responsive survey designs are strategies that inform adjustments in data collection procedures applied to sample units in the field based upon objective quantifiable metrics rather than ad hoc criteria

- **Improve representation of key subgroups**
- **Improve response rates in a meaningful manner**
- **Reduce design effects due to weighting**
- **Improve timeliness of survey estimates**
- **Reduce efforts on unproductive cases**

Overview of Metrics

- **Response rates**
 - Overall
 - Subgroups of interest
- **Stability of estimates as response increases**
- **Comparison of survey estimates to historical estimates for critical variables**
- **Measures of field effort**
- **Representativeness indicators**

Concepts of R-Indicators

R-indicators answer three questions

- 1. How well does the collected sample reflect the drawn sample (underlying population)?**
- 2. How much does each covariate contribute to the lack of sample representativeness?**
- 3. Which covariate category is causing lack of representativeness?**

R-Indicators

- **Sample R-Indicator**
 - Value between 0 and 1, with 1 being perfect sample representation
- **Unconditional Variable R-Indicator**
 - Value between 0.0 and 0.5, with 0 indicating the sample with respect to the associated variable has perfect representation
- **Unconditional Categorical R-Indicators**
 - Value between 0.0 and ± 0.5 , with 0 indicating the category of the associated variable has perfect representation

Building the Model

Development Process

- **Ensure that the portion of the sample we interview adequately represents the population of U.S. households (use ACS data for comparison)**
- **Identify Key Statistics for measuring representativeness**
 - Must use covariates available prior to the start of data collection
 - Keep variables consistent throughout field period
- **Use the 2016 SCF Area Probability (AP) to construct good logistic regression propensity models for the probability of response**
 - Allows us to build, simulate, and finalize a model using actual production data
 - Gives us historical information on sample performance and representativeness that can be compared to 2019 results

Variable Selection – Start with a full set of variables and compare output such as the adjusted R-squared and the Akaike information criterion (AIC) between models to determine final variable selection

Final Variable List:

- **Census Division**
- **Income below 200% Poverty**
- **Single Family Housing Units**
- **Owner Occupied Housing Units**
- **Bachelor's Degree Plus**
- **Black non-Latino**
- **White non-Latino**
- **Age 65 Plus**

Run R-Indicators on 2019 Data

Running the R-Indicators

- **Typically start to model when data collection hits about 50% of the target**
 - SCF began in April 2019
 - SCF hit approximately 50% of target completes in July of 2019
- **Start running R-Indicators on July data freeze**
 - Froze data once per month from July 2019 through February 2020
 - Ran full R-indicators for 8 different points in time (July through February)

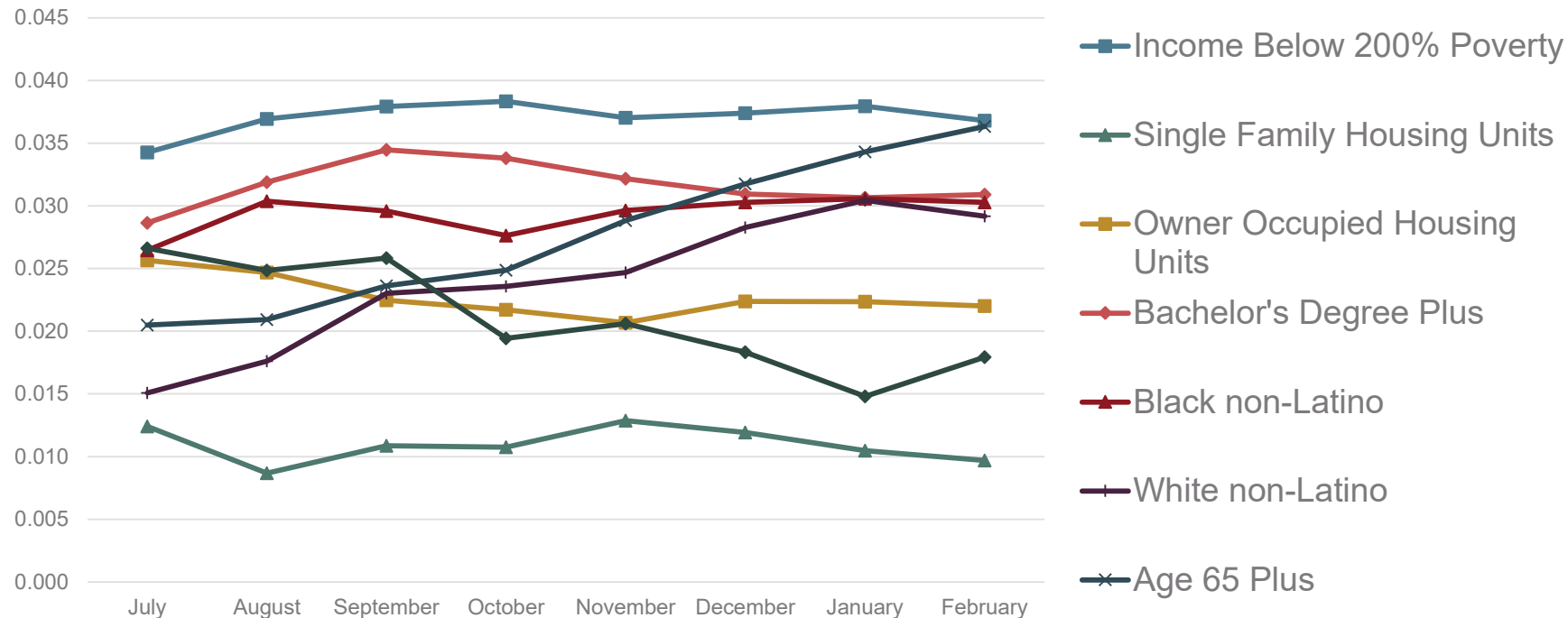
Full R-Indicator

- Ranges from 0 to 1, with a value of 1 indicating perfect sample representativeness of the underlying population
- Generally want Full R-Indicator above 0.75
- Better indicator is to compare with Full R-Indicator from previous rounds
- Full R-Indicator stayed relatively stable from middle to end of data collection
- Provides no direction as to where to concentrate collection resources

	2016	2019
Midpoint	0.804	0.814
August		0.814
September		0.810
October		0.813
November		0.814
December		0.813
January		0.815
Final	0.809	0.813

Partial R-Indicators

- Covariate R-indicators run from 0 to 0.5 with 0 being ideal
- Measures how representative the collected data are for key survey statistics
- Does not indicate which specific groups are under- or over-performing

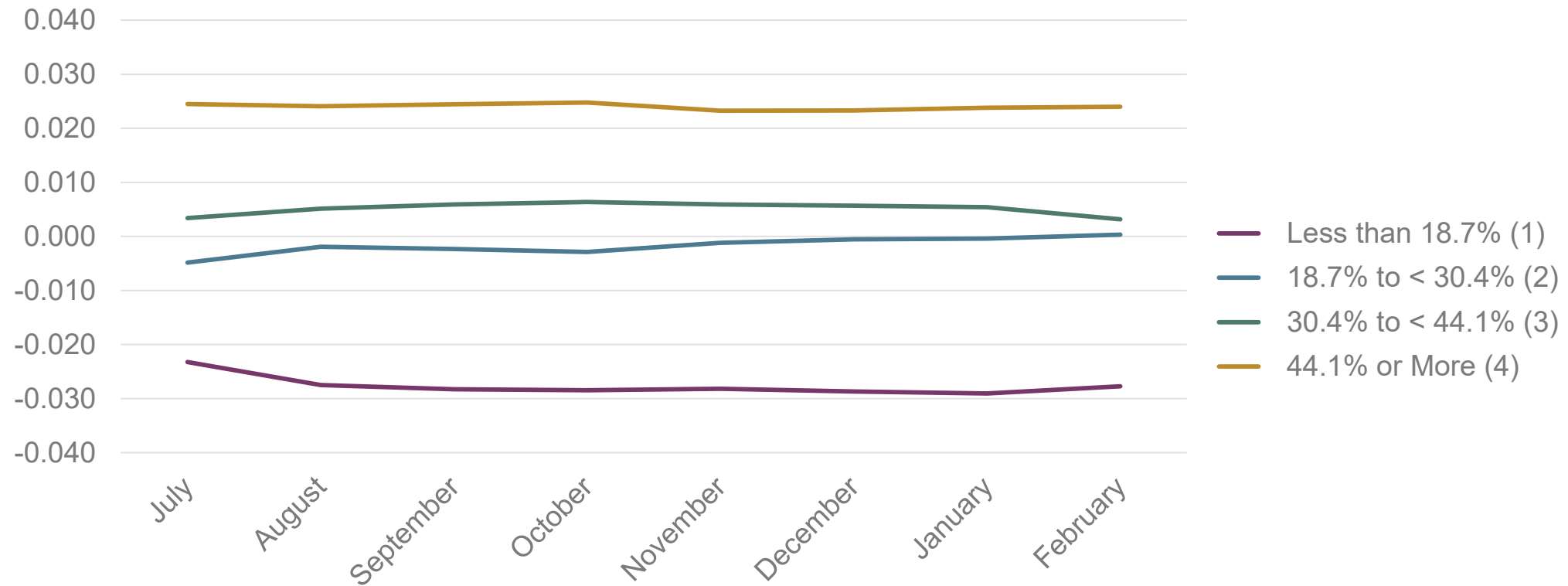


Category Partial R-Indicators

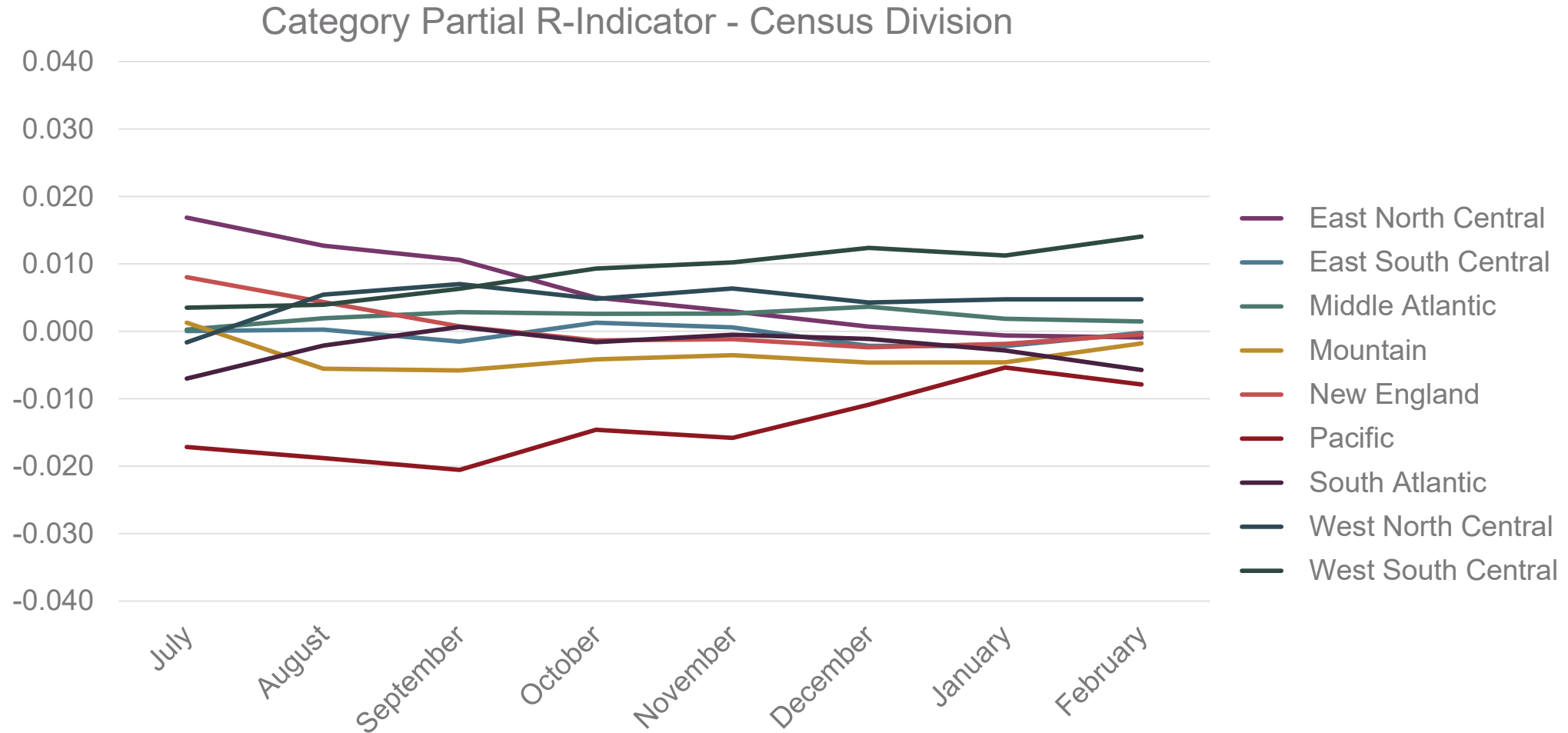
- Range from -0.5 to +0.5, with a value of 0 being ideal
- Shows which categories are over- or under-represented in the responding sample
- Positive values indicate this portion of the sample has better representativeness than those categories with a lower value

Category Partial R-Indicator Example #1

Category Partial R-Indicator - Income Below 200% Poverty



Category Partial R-Indicator Example #2



Results/Conclusions

- **Review early vs. late responders**
- **Use as a benchmark for overall sample representation**
- **Exploring use of R-indicators for future work**
 - Look for variables that historically have challenges in representation
 - Income Below 200% Poverty
 - Age 65 Plus
 - Bachelor's Degree Plus
 - Help inform future design
 - Increased outreach/incentives?

Katie Archambeau
Senior Research
Methodologist
archambeau-katie@norc.org

 Research You Can Trust™

 **NORC** at the
University of
Chicago

- Cohen, S. H., Lakhal, I., Resnick, D. (2018, July). *Tools for Implementing Adaptive Survey Designs at NORC*. Presented at the Joint Statistical Meetings, Vancouver.